



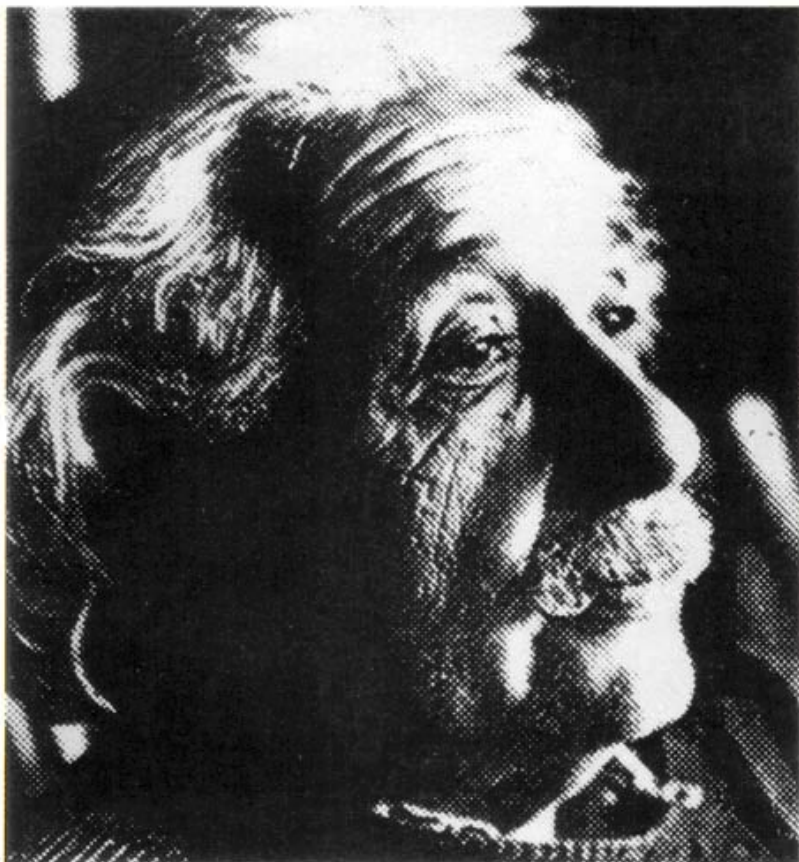
Albert - Einstein - Notas - Autobiograficas

Física

Universidade Estadual do Maranhão (UEMA)

53 pag.





Albert Einstein

Notas

Autobiográficas

4ª EDIÇÃO


EDITORA
NOVA
FRONTEIRA

Quando eu era um jovem razoavelmente precoce, fiquei impressionado com a futilidade das esperanças e dos esforços que atormentam incansavelmente os homens durante toda a sua vida. Além disso, muito cedo percebi a crueldade dessa busca, que naquele tempo era muito mais cuidadosamente disfarçada pela hipocrisia e por palavras brilhantes. Todos estavam condenados a participar dessa busca pela mera existência dos seus estômagos. O estômago talvez se saciasse com essa participação, mas não o homem, na medida em que é um ser pensante e dotado de sentimentos. A primeira válvula de escape era a religião, implantada nas crianças pela máquina educadora tradicional. Assim – embora fosse filho de pais absolutamente não-religiosos [judeus] –, entreguei-me a uma religiosidade profunda, que terminou abruptamente quando eu tinha apenas doze anos. A leitura de livros científicos populares convenceu-me de que a maioria das histórias da Bíblia não podia ser real. A consequência foi uma orgia positivamente fanática de livre-pensamento, combinada com a impressão de que a juventude é decididamente enganada pelo Estado com mentiras; foi uma descoberta esmagadora. Essa experiência fez com que eu passasse a desconfiar de todo tipo de autoridade, adotando uma atitude cética quanto às convicções vigentes em qualquer ambiente social específico – uma atitude que jamais abandonei, embora mais tarde tenha sido amenizada por uma visão mais perfeita das conexões causais.

A. Einstein.

Albert Einstein

Notas

Autobiográficas

Tradução de
Aulyde Soares Rodrigues

Título original
AUTOBIOGRAPHICAL NOTES – LIBRARY OF LIVING PHILOSOPHERS

© Library of Living Philosophers, and Estate of Albert Einstein
Direitos adquiridos para a língua portuguesa pela
EDITORA NOVA FRONTEIRA S.A.
Rua Bambina, 25 – CEP 22251 – Botafogo – Tel.: 286-7822
Endereço Telegráfico: NEOFRONT – Telex: 34695 ENFS BR – Rio de Janeiro, RJ

Capa
VICTOR BURTON

Revisão
JORGE AGUINALDO URANGA
LUIS AUGUSTO MESQUITA
CLARA RECHT DIAMENT

FICHA CATALOGRÁFICA
CIP-Brasil. Catalogação-na-fonte
Sindicato Nacional dos Editores de Livros, RJ.

Einstein, Albert.
E35n Notas autobiográficas / Albert Einstein. – Ed. comemorativa /
traduzida e anotada por Paul Arthur; tradução de Aulyde Soares Rodrigues.
Rio de Janeiro: Nova Fronteira, 1982.

Tradução de: Autobiographisches.

1. Einstein, Albert, 1879-1955. Título

82-0100

CDD-925
CDU-92Einstein

Apresenta a o

Em 1981, a Editora Nova Fronteira publicou, de Albert Einstein, Como vejo o mundo, onde se condensa seu pensamento humano e científico. O sucesso foi imediato e o livro se tornou um best-seller.

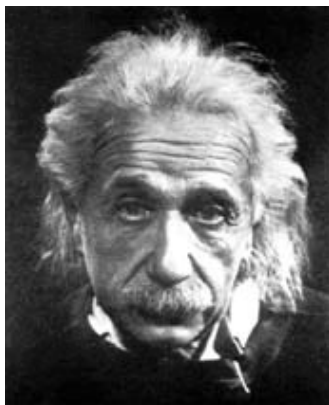
Agora, um ano depois, a mesma editora entrega aos leitores brasileiros estas Notas autobiográficas.

Nelas, Albert Einstein revela-se mais uma vez o excepcional expositor didático de suas teorias. Muito mais que uma autobiografia, trata-se de um retrato do desenvolvimento de sua mente e do modo como evoluiu sua visão científica dos problemas da física. Como todo gênio, ele não dá nenhuma importância à sua vida pessoal e não faz nenhuma referência a ela.

Tendo descoberto desde cedo que os jovens sempre são enganados pelo Estado em suas mais legítimas aspirações, Einstein logo assumiu como comportamento básico a desconfiança de todo tipo de autoridade, adotando uma atitude crítica quanto às convicções vigentes em qualquer ambiente social específico.

esse espírito essencialmente lúcido, sempre voltado para a interrogação crítica, marcado pelo destemor e pela coerência numa liberdade que a nada se submeteu, é esse exemplo de integridade absoluta que o leitor brasileiro encontrará nestas Notas.

Não melhor do que elas e compreender sua motivação interior, sobretudo numa época em que não poucos, em função do poder e da glória, abdicam de seus princípios e de seu saber.



Albert Einstein

(Foto de Philippe Halsman)

Prefácio do Editor Norte-Americano

As Notas Autobiográficas de Albert Einstein, já falecido, constituem um documento raro e precioso. Foi a única e importante tentativa do professor Einstein no sentido de escrever algo que se aproximasse da forma autobiográfica.¹ Para ele, consistiu apenas em uma exposição sobre o desenvolvimento de sua mente e de como uma série de pensamentos e considerações levaram a outros; resumindo, um relato de como e por que passou a pensar de certa forma e quais as conclusões a que foi levado, em um determinado momento. Embora seja uma narrativa eminentemente pessoal, quase nada esclarece sobre sua vida privada ou familiar, ou sobre os tremendos acontecimentos que abalaram o mundo e que envolveram sua existência quotidiana. Em outras palavras, é um Selbst-Darstellung (auto-retrato) científico do mais original pensador do século vinte.

O livro foi escrito para atender ao insistente pedido do editor e posso garantir que foi preciso usar de muita persuasão para conseguir o volume VII da nossa Biblioteca de Filósofos Vivos, intitulado Albert Einstein: Filósofo-Cientista (publicado pela primeira vez em 1949). A partir de 1949, foi publicado em inglês (e até mesmo no original alemão) apenas nas várias edições daquele volume. Estão sendo lançados agora nas duas línguas, uma ao lado da outra pela primeira vez como um volume separado, em

¹ A exceção é o “Autobiographische Skizze”, menos importante, com oito páginas, publicado no *Helle Zeit-Dunkle Zeit*, in *Memoriam Albert Einstein* de Carl Selig. (Europa Verlag, Zurique, 1956, pp. 9-17).

comemora a o ao centenário do nascimento de Einstein, em 14 de maio de 1879.

A tradução inglesa, originalmente feita pelo editor, foi beneficiada com uma leitura minuciosa e revisão (quando necessária) feitas pelo professor Peter Bergmann, físico de renome da Universidade de Siracusa, que durante cinco anos trabalhou como assistente científico do Dr. Einstein no Instituto de Estudos Avançados de Princeton. O professor Bergmann e os administradores do patrimônio de Einstein, Dr. Otto Nathan e Srta. Helen Dukas, prestaram colaboração extremamente cortês e valiosa que o editor tem satisfação de agradecer.

Na verdade, graças ao Dr. Nathan pudemos reproduzir, no frontispício deste livro, a bela e interessante fotografia feita pelo senhor Philippe Halsman.

Agradecemos também Fundação Hegeler e aos administradores da Open Court Publishing Company, de La Salle, Illinois, que conseguiram, quase em tempo recorde, produzir este livro em formato especial comemorativo, para ser publicado por ocasião do centenário do nascimento de Einstein, uma data que a Southern Illinois University, em Carbondale, tem o prazer de comemorar com a Semana Einstein especial, de 23 de fevereiro a 3 de março de 1979.

Paul Arthur Schilpp
Carbondale, Illinois
Junho, 1978

Notas Autobiográficas

Aos sessenta e sete anos de idade, preparo-me para escrever algo que é como o meu obituário. Não o faço unicamente devido à insistência do Schilpp, mas porque na verdade acredito que é válido mostrar àqueles que lutam ao nosso lado uma retrospectiva da nossa própria luta e das nossas pesquisas. Após alguma reflexão, concluí que essa tentativa provavelmente será imperfeita. Pois, por mais breve e limitada que seja a carreira de um homem, e por maior que seja o índice de erro possível, a exposição de tudo aquilo que é digno de ser comunicado não é fácil tarefa – um homem com sessenta e sete anos não é de modo nenhum o mesmo homem que era aos 50, 30 ou 20. Todas as reminiscências são coloridas com os tons do presente, vistas portanto sob uma falsa perspectiva. Essa consideração poderia ser suficiente para me deter. Contudo, há muita coisa na nossa experiência que não é evidente ao pensamento de muitos.

Quando eu era um jovem razoavelmente precoce, fiquei impressionado com a futilidade das esperanças e dos esforços que atormentam incansavelmente os homens durante toda a sua vida. Além disso, muito cedo percebi a crueldade dessa busca, que naquele tempo era muito mais cuidadosamente disfarçada pela hipocrisia e por palavras brilhantes. Todos estavam condenados a participar dessa busca pela mera existência dos seus estômagos. O estômago talvez se saciasse com essa participação, mas não o homem, na medida em que é um ser pensante e dotado de sentimentos. A primeira válvula de escape era a religião, implantada nas crianças pela máquina educadora tradicional. Assim – embora fosse filho de pais absolutamente não-religiosos [judeus] –, entreguei-me a uma religiosidade profunda, que terminou abruptamente quando tinha

apenas doze anos. A leitura de livros científicos populares convenceu-me de que a maioria das histórias da Bíblia não podia ser real. A consequência foi uma orgia positivamente fanática de livre-pensamento, combinada com a impressão de que a juventude é decididamente enganada pelo Estado, com mentiras; foi uma descoberta esmagadora. Essa experiência fez com que passasse a desconfiar de todo tipo de autoridade, adotando uma atitude cética quanto às convicções vigentes em qualquer ambiente social específico – uma atitude que jamais abandonei, embora mais tarde tenha sido amenizada por uma visão mais perfeita das conexões causais.

É evidente para mim que o paraíso religioso da juventude, assim perdido, foi a primeira tentativa para me libertar das cadeias do “mero individualismo”, de uma existência dominada por desejos, esperanças e sentimentos primários. Além de mim, fora de mim, estava o mundo imenso, que existe independente dos seres humanos e que se nos apresenta como um enorme e eterno enigma, em parte acessível à nossa observação e ao nosso pensamento. A contemplação desse mundo acenava-me como uma força libertadora, e percebi que muitos daqueles a quem aprendera a respeitar e admirar haviam encontrado, por esse meio, a liberdade interior e a segurança. A conquista mental desse mundo extra-individual dentro dos limites da capacidade humana se me apresentava meio consciente e meio inconscientemente como o objetivo supremo. Os homens do presente e do passado, similarmente motivados, bem como os conhecimentos que tinham adquirido, eram dignos de serem cultivados como amigos. O caminho para esse paraíso não era tão confortável e tão atraente quanto o que levava ao paraíso religioso; mas provou ser digno de confiança e jamais me arrependi de tê-lo escolhido.

O que afirmei acima é verdadeiro apenas em certo sentido, do mesmo modo que algumas linhas de um esboço mostram um objeto complicado, cheio de detalhes, apenas até certo ponto. Se um indivíduo cultiva pensamentos ordenados, é bem possível que essa parte da sua natureza venha a se tornar mais pronunciada à custa das outras partes, determinando assim a sua mentalidade. Neste caso, pode acontecer que o indivíduo veja, em retrospecto, um

desenvolvimento sistematicamente uniforme, enquanto que as experiências atuais se processam em situações caleidoscópicas especiais. A grande variedade das situações externas e o estreito campo do conteúdo momentâneo de conhecimento consciente criam uma espécie de atomização da vida de todo ser humano. Para um homem do meu tipo, o ponto decisivo do desenvolvimento encontra-se no desengajamento gradual do centro de interesse para um grau muito além do momentâneo e do puramente individual, voltando-se para a conquista de uma compreensão conceptual das coisas. Sob esse ponto de vista, as afirmações esquemáticas feitas acima contêm toda a verdade que pode ser exposta do modo mais breve.

O que, exatamente, é o pensamento? Quando, na percepção das impressões sensoriais, emergem figuras da memória, isto ainda não é “pensar”. E quando esses quadros formam seqüências, cada membro criando o outro, isto também ainda não é “pensar”. Porém, quando uma certa figura aparece em várias seqüências, nesse caso – precisamente devido a essa recorrência – torna-se um elemento de organização para tais seqüências, no sentido de unir seqüências que por si mesmas não se relacionam entre si. Esse elemento vem a ser um instrumento, um conceito. Creio que a transição da livre associação ou “sonho” para o pensamento caracteriza-se pelo papel mais ou menos importante representado pelo conceito. Não é de modo algum necessário que o conceito esteja ligado a um signo que possa ser reconhecido e reproduzido pelos sentidos (palavra), mas, quando isto se dá, o pensamento torna-se, por esse meio, capaz de ser comunicado.

Com que direito – perguntará o leitor – o homem opera com tal descuido e de forma tão elementar com idéias, nesse reino tão problemático, sem ao menos tentar provar alguma coisa? Minha defesa: todos os nossos pensamentos têm a natureza do jogo livre dos conceitos; a justificativa desse jogo está no grau de compreensão das sensações que podemos alcançar com a sua ajuda. O conceito de “verdade” não pode ainda ser aplicado a essa estrutura; na minha opinião, esse conceito só é aplicável quando temos à mão um acordo (*convenção*) que abrange os elementos e as regras do jogo.

Não tenho dúvidas de que o nosso pensamento se processa, na maior parte das vezes, sem o uso dos signos (palavras) e, além disso, em grande parte inconscientemente. Se assim não fosse, como seria possível “lembrarmos com estranheza” e de forma espontânea uma determinada experiência? Essa “lembança inquisitiva” pode ocorrer quando a experiência está em conflito com conceitos bem estabelecidos em nossa mente. Sempre que experimentamos esse conflito aguda e intensamente, ele reage contra nosso mundo mental de modo decisivo. O desenvolvimento desse mundo mental é, em certo sentido, uma fuga constante do “pensamento de estranheza”.

Aos 4 ou 5 anos, experimentei esse sentimento quando meu pai mostrou-me uma bússola. O fato de a agulha comportar-se de uma certa forma não se encaixava entre os tipos de ocorrências que podiam ser colocados no mundo inconsciente dos conceitos (eficácia produzida pelo “toque” direto). Lembro-me ainda – ou pelo menos creio que me lembro – que essa experiência causou-me uma impressão profunda e duradoura. Devia haver algo escondido nas profundezas das coisas. Aquilo que o homem conhece desde a infância não provoca esse tipo de reação; não se surpreende com o vento e a chuva, com a lua, nem com o fato de essa mesma lua não cair do céu, ou com as diferenças entre a matéria viva e a matéria sem vida.

Aos doze anos experimentei minha segunda sensação de espanto, de natureza completamente diversa da primeira, provocada por um livrinho de geometria plana de Euclides, que veio ter às minhas mãos no início do ano escolar. Ali estavam afirmações como, por exemplo, a intersecção das três alturas do triângulo num determinado ponto que – embora não fosse evidente – podia ser provada com tal certeza que qualquer dúvida estava fora de cogitação. Esta certeza lúcida impressionou-me profundamente. O fato de os axiomas serem aceitos sem prova não me perturbou. De qualquer forma, era bastante poder basear as provas em proposições cuja validade me parecia livre de qualquer dúvida. Por exemplo, lembro-me que um tio me falou sobre o teorema de Pitágoras antes que eu tivesse lido o livrinho sagrado de geometria. Com muito esforço consegui “provar” esse teorema, tomando como base a similaridade dos triângulos; parecia-me “evidente” que as relações

dos lados dos triângulos de ângulos retos teriam de ser completamente determinadas por um dos ângulos agudos. Para mim, apenas as idéias que não eram evidentes dessa forma precisavam ser provadas. Além disso, os objetos tratados pela geometria não me pareciam diferentes dos objetos da percepção sensorial “que podem ser vistos e tocados”. Esse conceito primário, que provavelmente está no fundo da conhecida crítica de Kant sobre a possibilidade de “julgamentos sintéticos *a priori*”, repousa obviamente no fato de que a relação dos conceitos geométricos com os objetos da experiência direta (barra rígida, intervalo finito etc.) existia no inconsciente.

Assim, se aparentemente é possível chegar-se a um conhecimento dos objetos da experiência por meio do pensamento puro, essa “estranheza” tinha como base o erro. Contudo, para quem a experimenta pela primeira vez, parece maravilhoso o homem ser capaz de alcançar tal grau de certeza e de pureza de pensamento, como nos demonstraram os gregos com sua geometria.

Agora que tomei um desvio, interrompendo o meu obituário apenas iniciado, não hesitarei em apresentar em poucas palavras meu credo epistemológico, embora já tenha dito algo sobre o mesmo nas considerações acima expostas. Na verdade, esse credo desenvolveu-se muito mais tarde, e lentamente, e não corresponde ao meu modo de pensar quando era jovem.

Vejo de um lado a totalidade das experiências sensoriais e, do outro, a totalidade dos conceitos e proposições descritos nos livros. As relações entre os conceitos e as proposições são de natureza lógica e o processo do pensamento lógico é estritamente limitado à efetivação da conexão entre os conceitos e as proposições entre si, de acordo com regras firmemente estabelecidas, que constituem a matéria da lógica. Os conceitos e proposições adquirem “sentido” ou “conteúdo” apenas através das suas conexões com as experiências sensoriais. A conexão destas últimas com os primeiros é puramente intuitiva, e não de natureza lógica em si mesma. O grau de certeza com o qual essa conexão ou ligação intuitiva pode ser admitida é a única diferença entre a fantasia desprovida de conteúdo e a “verdade” científica. O sistema de conceitos é criação do homem, bem como as regras da sintaxe, que constituem a estrutura dos sistemas

conceptuais. Embora esses sistemas sejam logicamente arbitrários na sua totalidade, são restritos pelo objetivo de permitir a mais completa e correta coordenação (intuitiva) com a totalidade das experiências sensoriais; em segundo lugar, objetivam a maior escassez possível dos seus elementos logicamente independentes (conceitos básicos e axiomas), isto é, seus conceitos definidos e proposições não-derivadas [postuladas].

Uma proposição é correta quando, dentro de um sistema lógico, é deduzida de acordo com as regras aceitas da lógica. Um sistema tem conteúdo de verdade de acordo com a certeza e a inteireza da possibilidade de coordenação com a totalidade da experiência. Uma proposição correta tem a sua “verdade” adquirida por empréstimo ao conteúdo de verdade do sistema a que pertence.

Uma observação sobre o desenvolvimento histórico. Hume compreendeu claramente que certos conceitos, como por exemplo o da causalidade, não podem ser deduzidos do material da experiência por meio de métodos lógicos. Kant, convencido da necessidade essencial de certos conceitos, considerou-os – do modo como são selecionados – como premissas necessárias para qualquer tipo de pensamento, e fez a distinção entre esses conceitos e os conceitos de origem empírica. Porém estou convencido de que essa distinção está errada, ou pelo menos não faz justiça ao problema de modo natural. Todos os conceitos, mesmo os que estão mais próximos da experiência, são, do ponto de vista da lógica, postulados livremente escolhidos, como o conceito da causalidade, ponto de partida dessa crítica.

Voltemos agora ao obituário. Dos doze aos dezesseis anos, familiarizei-me com os elementos da matemática, incluindo os princípios do cálculo diferencial e cálculo integral. Tive a sorte de encontrar livros que não se preocupavam com o rigor lógico, mas que permitiam a apresentação clara das idéias principais. Era um trabalho verdadeiramente fascinante; certos pontos extremos me impressionavam tanto quanto os da geometria elementar – a idéia básica da geometria analítica, as séries infinitas, os conceitos de derivadas e de integrais. Tive a sorte também de aprender os resultados essenciais e os métodos de todo o campo das ciências

naturais, numa excelente obra popular que se limitava quase que exclusivamente aos aspectos qualitativos (Bernstein, *Popular Books on Natural Science*, em cinco ou seis volumes), e que li com absorvente atenção. Já estudara também um pouco de física teórica quando, com dezessete anos, entrei para o Instituto Politécnico de Zurique para estudar matemática e física.

No Instituto tive ótimos professores (por exemplo, Hurwitz, Minkowski) e aprendi matemática a fundo. Trabalhei a maior parte do tempo no laboratório de física, fascinado pelo contato direto com a experiência. O resto do tempo era quase todo utilizado estudando em casa as obras de Kirchhoff, Helmholtz, Hertz, etc. O fato de ter negligenciado a matemática até certo ponto deve-se não apenas ao meu maior interesse pelas ciências naturais, mas também a uma experiência singular. A matemática dividia-se em numerosas especializações, cada uma delas capaz de absorver o curto tempo de vida que nos é dado. Portanto, vi-me na posição o asno de Buridan², incapaz de se decidir entre vários montes de feno. Talvez minha intuição não fosse tão desenvolvida no campo da matemática a ponto de diferenciar com clareza o que era fundamentalmente importante, realmente básico, do resto da erudição mais ou menos dispensável. Além disso, meu interesse pelo estudo da natureza era sem dúvida mais forte, e não estava ainda bem claro para mim, apenas um jovem estudante, o fato de que o acesso ao conhecimento mais profundo dos princípios básicos da física depende dos métodos matemáticos mais complexos. Só vim a reconhecer esse fato gradualmente, depois de anos de trabalho científico independente. É verdade que a física

² Buridan (João), doutor escolástico do século XIV, n. em Béthune, França; m. depois de 1358. Seu nome ficou ligado ao famoso argumento do *asno de Buridan*. Buridan colocava a hipótese de um asno atormentado com igual intensidade pela fome e pela sede, que se encontrava a igual distância de um balde de água e de uma ração de feno. Por onde começaria para satisfazer as duas necessidades igualmente prementes? É o problema da *liberdade de indiferença*, que significa a situação de um homem perplexo entre duas resoluções que o tentem por igual. Cumpre notar que não se encontra vestígio desse argumento nos escritos de Buridan. Segundo uma lenda muito contestável, Buridan teria sido cúmplice de Margarida de Borgonha nas suas proezas na Torre de Nesles. (N. do T.)

também está dividida em vários campos, cada um deles capaz de devorar uma curta vida de trabalho sem satisfazer a fome de um conhecimento mais profundo. Aqui também era enorme a escassez de dados experimentais referentes à matéria. Entretanto, nesse campo logo aprendi a reconhecer os caminhos que conduziam às noções fundamentais, deixando de lado todo o resto, tudo aquilo que sobrecarrega a mente desviando-a do essencial. O problema era que, como estudantes, éramos obrigados a acumular essas noções em nossas mentes para os exames. Esse tipo de coerção tinha (para mim) um efeito frustrante. Depois de ter passado nos exames finais, passei um ano inteiro durante o qual qualquer consideração sobre problemas científicos me era extremamente desagradável. Porém, devo dizer que na Suíça essa coerção era bem mais branda do que em outros países, onde a verdadeira criação científica é completamente sufocada. Fazíamos apenas dois exames; durante o resto do tempo podíamos nos dedicar ao que bem entendêssemos. Isso se dava especialmente quando se tinha um amigo, como era o meu caso, que assistia às aulas regularmente e anotava a matéria. Assim eu ficava livre para qualquer atividade, até alguns meses antes do exame, uma liberdade que aproveitei ao máximo, assumindo alegremente o peso na consciência como o menor entre dois males. Na verdade, é quase um milagre que os métodos modernos de instrução não tenham exterminado completamente a sagrada sede de saber, pois essa planta frágil da curiosidade científica necessita, além de estímulo, especialmente de liberdade; sem ela, fenece e morre. É um grave erro supor que a satisfação de observar e pesquisar pode ser promovida por meio da coerção e da noção do dever. Muito ao contrário, acredito que seria possível eliminar por completo a voracidade de um animal predatório obrigando-o, à força, a se alimentar continuamente, mesmo quando não tem fome, especialmente se o alimento usado para a coerção for escolhido para isso.

Descreverei agora a situação do campo da física naquela época. Apesar da grande produtividade em casos isolados, prevalecia uma

rigidez dogmática em relação aos princípios. No começo (se é que houve um começo), Deus criou as leis do movimento, de Newton, junto com as massas e as forças. Isso era tudo. Além dessa verdade, segue-se, por dedução, o desenvolvimento do método matemático adequado. O que foi conseguido no século dezenove, com base nesse princípio, especialmente através da aplicação das equações diferenciais parciais, merece a admiração de todos os que podem compreender a importância de tais atos. Provavelmente, Newton foi o primeiro a revelar, na sua teoria da propagação do som, a eficácia das equações diferenciais parciais. Euler criara os fundamentos da hidrodinâmica. O desenvolvimento mais sofisticado da mecânica das massas distintas, como base de toda a física, foi a grande conquista do século dezenove. Entretanto, o que mais me impressionou quando estudante não foi tanto o desenvolvimento da mecânica ou a solução de problemas complexos, mas as conquistas da mecânica em áreas aparentemente não ligadas a essa parte da física. A teoria mecânica da luz, que define a luz como o movimento de onda de um éter elástico semi-rígido; e, acima de tudo, a teoria cinética dos gases; a independência do calor específico dos gases não-atômicos; do peso atômico; a derivação da equação do estado de um gás e sua relação com o calor específico; a teoria cinética da dissociação dos gases e a importante relação quantitativa entre viscosidade, condução e calor e difusão dos gases, que deu a medida da grandeza absoluta do átomo. Esses resultados provaram que a mecânica é a base da física e das hipóteses atômicas, mais tarde implantadas definitivamente na química. Porém, na química, apenas as razões das massas atômicas desempenhavam um papel, e não a sua grandeza absoluta. Portanto, a teoria atômica podia ser encarada mais como um símbolo visual do que como conhecimento sobre a composição real da matéria. Além disso, foi de grande importância o fato de a teoria estatística da mecânica clássica ter conseguido deduzir as leis básicas da termodinâmica, o que de certa forma já fora realizado, em essência, por Boltzmann.

Não nos deve portanto surpreender o fato de todos os físicos do século passado terem encarado a mecânica clássica como uma fundação sólida e definitiva de toda a física – na verdade, de toda a

ciência natural – e que tenham insistido na tentativa de basear a teoria do eletromagnetismo de Maxwell, que nessa época começava lentamente a ser reconhecida, também na mecânica. Maxwell e Hertz, que podem ser considerados os cientistas que abalaram a crença de que a mecânica é a base final e definitiva do pensamento físico, eles também, no seu pensamento consciente, atinham-se à noção de que a mecânica era a base da física. Foi Ernst Mach quem, com sua *History of Mechanics*, revolucionou essa crença dogmática. Esse livro exerceu em mim uma influência profunda quando era ainda estudante. Vejo a grandeza de Mach no seu ceticismo e incorruptível independência; contudo, na minha juventude, a posição epistemológica de Mach influenciou-me acentuadamente, uma posição que hoje considero impossível de ser mantida. Pois ele não focalizou devidamente a natureza essencialmente construtiva e especulativa de todo pensamento e principalmente do pensamento científico. Em consequência, condenou a teoria precisamente nos pontos em que esse caráter construtivo-especulativo é claro e indiscutível, como na teoria cinética dos átomos.

Antes de iniciar a crítica da mecânica como base fundamental da física, devo abordar algumas generalidades sobre o ponto de vista crítico que pode ser aplicado às teorias físicas. O primeiro é óbvio: a teoria não deve contradizer os fatos empíricos. Por mais evidente que seja essa exigência, sua aplicação é bastante complexa. Pois é possível, muitas vezes, talvez até sempre, conservar um fundamento teórico geral adaptando-o aos fatos, adicionando-se pressupostos artificiais. Entretanto, em qualquer caso, este primeiro ponto de vista refere-se à confirmação do fundamento teórico por meio dos fatos empíricos disponíveis.

O segundo ponto de vista não diz respeito ao relacionamento com a observação, mas com as premissas da própria teoria, ou aquilo que pode ser descrito como a “naturalidade” ou “simplicidade lógica” das premissas (os conceitos básicos e sua relação mútua). Este ponto de vista, cuja formulação exata apresenta grandes dificuldades, desempenhou papel importante na seleção e avaliação das teorias desde tempos imemoriais. O problema não se restringe a uma espécie de enumeração das premissas logicamente

independentes (se isso fosse possível sem ambigüidade), mas a um tipo de avaliação recíproca das qualidades incomensuráveis. Além disso, entre as teorias com bases igualmente “simples”, deve ser considerada superior aquela que delimita de modo mais definido as qualidades dos sistemas, que, de outro modo, seriam impraticáveis (isto é, contêm afirmações mais específicas). Não preciso citar a “abrangência” das teorias, uma vez que estamos nos limitando às que têm como objeto a *totalidade* dos fenômenos físicos. O segundo ponto de vista pode ser definido brevemente como relacionado à “perfeição interna” da teoria, enquanto o primeiro refere-se à confirmação externa da mesma. Considero também como parte da “perfeição interna” da teoria o seguinte: damos mais valor a uma teoria quando, sob o ponto de vista lógico, ela não implica uma escolha arbitrária entre teorias equivalentes e que possuem estruturas análogas.

Não tentarei desculpar a falta de precisão das afirmativas contidas nos dois últimos parágrafos, alegando escassez de espaço à minha disposição; devo confessar que não posso, neste ponto, e talvez de modo algum, substituir essas sugestões por definições mais precisas. Acredito, entretanto, ser possível uma formulação mais definida. De qualquer modo, acontece que entre os “oráculos” há um consenso geral no sentido de julgar a “perfeição interna” das teorias e, especialmente, no que se refere ao grau de “confirmação externa”.

Passo agora à crítica da mecânica como base da física.

Quanto ao primeiro ponto de vista (confirmação pela experiência), a incorporação das ondas ópticas no quadro da mecânica estava destinada a criar sérias dúvidas. Se a luz era interpretada como um movimento ondulatório num corpo elástico (éter), este devia ser um meio capaz de impregnar tudo, devido à transversalidade das ondas luminosas, assemelhando-se, de um modo geral, a um corpo sólido mas compressível, de modo que as ondas longitudinais não existissem. Esse éter deveria existir como um fantasma ao lado da matéria, uma vez que aparentemente não ofereceria resistência alguma ao movimento dos corpos “ponderáveis”. Para explicar os índices de refração dos corpos transparentes, bem como o processo de emissão e absorção da

radiação, é preciso pressupor interações complexas entre os dois tipos de matéria, algo que não foi de modo algum tentado seriamente, e muito menos alcançado.

Além disso, as forças eletromagnéticas necessitam da introdução de massas elétricas que, embora não possuam inércia aparente, atuam entre si, e essa interação é, em contraste com a força de gravitação, do tipo polar.

O que levou finalmente os físicos, após longa hesitação, a abandonar a crença na possibilidade de toda a física ter como base a mecânica de Newton foi a eletrodinâmica de Faraday e Maxwell. Essa teoria, confirmada pelas experiências de Hertz, provou a existência de fenômenos eletromagnéticos que por sua própria natureza são separados de toda matéria ponderável a saber, as ondas no espaço vazio, que consistem em “campos” eletromagnéticos. Para que a mecânica fosse mantida como base fundamental da física, a equação de Maxwell precisaria ser interpretada mecanicamente. Isso foi tentado arduamente, mas sem resultado, ao passo que as equações em si mesmas tornaram-se cada vez mais frutíferas. O cientista habituava-se a operar com esses campos como se fossem substâncias independentes, sem necessidade de levar em conta sua natureza mecânica; assim, a mecânica, como base da física, ia sendo abandonada quase imperceptivelmente, porque sua adaptação aos fatos representava uma tarefa inútil. A partir daí, passaram a existir dois tipos de elementos conceptuais: de um lado, pontos materiais com forças a certa distância entre eles e, do outro lado, o campo contínuo. Estamos agora em um estágio intermediário da física sem uma base uniforme para o todo, um estágio que – embora não-satisfatório – está muito longe de ser ultrapassado.

Agora algumas observações sobre a crítica da mecânica como base fundamental da física, de acordo com o segundo ponto de vista, o do “interior”. No atual estado da ciência, isto é, depois do abandono da mecânica como fundamento da física, essa crítica tem uma importância apenas metodológica. Porém, é pertinente no

sentido de mostrar o tipo de argumentação que, na seleção das teorias do futuro, desempenhará um papel cada vez mais importante, à medida que os conceitos mais básicos e os axiomas forem afastados de tudo o que pode ser observado diretamente, de modo que o confronto das implicações das teorias com os fatos se torne cada vez mais difícil e mais remoto. O primeiro da lista a ser mencionado é o argumento de Mach que, incidentalmente, já fora reorganizado por Newton de modo claro (a experiência com o balde). Sob o ponto de vista da pura descrição geométrica, todos os sistemas “rígidos” de coordenadas são logicamente equivalentes. As equações da mecânica (por exemplo, a lei da inércia) reivindicam validade apenas no que se refere a uma classe específica desses sistemas, isto é, os “sistemas de inércia”. Nesse contexto, o sistema coordenado não tem significado como objeto material. Portanto, para justificar a necessidade dessa escolha específica é preciso que se encontre algo que exista além dos objetos (massas, distâncias) de que trata a teoria. Por esse motivo, o “espaço absoluto” como determinante original foi introduzido por Newton como participante ativo onipresente em todos os fenômenos mecânicos; obviamente, o “absoluto” era definido como influenciado pelas massas e pelo movimento. O que transforma este estado de coisas em algo extremamente desagradável é o pressuposto da existência de vários sistemas de inércia relacionados entre si, numa transferência uniforme e irracional, e que supostamente devem ser distinguidos entre todos os outros sistemas rígidos.

Mach argumenta que uma teoria razoável da inércia dependeria da interação das massas, exatamente como é verdadeiro para as outras forças de Newton, uma idéia que durante muito tempo considerei correta. Entretanto, pressupõe implicitamente que a teoria básica deveria ser, de um modo geral, do tipo da mecânica de Newton: as massas e suas interações como conceitos originais. Essa tentativa de solução não se ajusta a uma teoria consistente do campo, como se pode perceber imediatamente.

Entretanto, a lógica da crítica de Mach pode ser claramente demonstrada com o seguinte exemplo. Imaginemos estudiosos da mecânica que conheçam apenas uma pequena parte da superfície terrestre e que não podem avistar as estrelas. Sua tendência será de

atribuir certas qualidades físicas à dimensão vertical do espaço (direção da aceleração dos corpos que caem) e, baseados nesse conceito, apresentarão provas de que a terra é horizontal em sua maior parte. Podem não se deixar influenciar pelo argumento de que o espaço, nas suas propriedades geométricas, é isotrópico e que, portanto, seria insatisfatório postular leis básicas da física segundo as quais existe uma direção preferencial; provavelmente afirmarão (como Newton) o vertical absoluto, provado pela experiência, como algo com que devemos nos acostumar. A preferência pela verticalidade sobre todas as outras direções espaciais é análoga à preferência dada aos sistemas de inércia sobre os outros sistemas coordenados rígidos.

Consideremos agora outros argumentos que tratam também da simplicidade interna, ou naturalidade da mecânica. Se aceitarmos os conceitos de espaço (incluindo a geometria) e de tempo, sem nenhuma questão de natureza crítica, não haverá razão para objeções à idéia de ação a distância, embora tal conceito não se adapte às idéias que temos, baseadas na experiência comum. Entretanto, outra consideração faz com que a mecânica, tomada como base fundamental da física, pareça primária.

Existem duas leis essenciais:

- (1) a lei do movimento
- (2) a expressão para a força da energia potencial.

A lei do movimento é precisa, embora vazia enquanto não é dada a expressão para as forças. Entretanto, para postular essa expressão existe uma extensa gama de arbitrariedade, especialmente se ignorarmos o requisito, que de qualquer forma não é muito natural, de que as forças devem depender apenas das coordenadas (e não, por exemplo, de suas derivadas, com respeito ao tempo). Somente dentro da teoria é completamente arbitrário o fato de serem as forças gravitacionais (e a eletricidade), que partem de um ponto, governadas pela função potencial ($1/r$). Observação adicional: há muito se sabe que essa função é a solução esfericamente simétrica da mais simples (rotação-invariante) equação diferencial! $\nabla^2 \phi = 0$; não

seria portanto absurdo considerar isto como um sinal de que essa função deve ser julgada resultante da lei espacial, uma abordagem que eliminaria a arbitrariedade na lei da força. Esta é, na verdade, a primeira indicação sugestiva de um afastamento da teoria da ação a distância que na realidade só se iniciou mais tarde preparado por Faraday, Maxwell e Hertz em resposta à pressão externa dos dados experimentais.

Gostaria de mencionar ainda uma assimetria interna dessa teoria: a massa inerte que ocorre na lei do movimento aparece também na lei da força gravitacional, mas não nas expressões das outras forças. Finalmente, é preciso acentuar o fato de que a divisão da energia em energia cinética e energia potencial deve ser considerada como não-natural; H. Hertz considerou esse fato tão estranho que, no seu último trabalho, tentou libertar a mecânica do conceito da energia potencial (isto é, do conceito de força).

É o suficiente sobre esse assunto. Perdoe-me, Newton; você descobriu talvez o único caminho possível em sua época para um homem possuidor do mais alto raciocínio e poder criativo. Os conceitos que criou ainda hoje orientam o nosso pensamento na física, embora saibamos que deverão ser substituídos por outros, muito afastados da esfera da experiência imediata, para possibilitar a compreensão mais profunda dos relacionamentos.

“Isto é um obituário?”, deverá estar indagando o leitor atônito. Permita-me que responda: em essência, sim. Pois, para um ser humano do meu tipo, o essencial está precisamente *naquilo* que pensa e *como* pensa, e não nas coisas que faz ou que lhe são feitas. Portanto, um obituário pode limitar-se, em sua maior parte, à comunicação dos pensamentos que desempenharam um papel considerável na minha luta. Quanto maior a simplicidade das premissas, mais impressionante é a teoria, maior o número de coisas diferentes com as quais se relaciona e mais extensa sua área de aplicação. Daí a profunda impressão que me causou o conhecimento da termodinâmica clássica. É a única teoria física de conteúdo universal que,

estou convencido, dentro da estrutura da aplicabilidade dos seus conceitos básicos, jamais será derrubada (para a atenção especial dos céticos por princípio).

Quando eu estava ainda na universidade, o assunto mais fascinante era a teoria de Maxwell. O que lhe dava um aspecto revolucionário era a transição da ação a distância para os campos, como variáveis fundamentais. A incorporação da óptica à teoria do eletromagnetismo, ao lado da relação entre a velocidade da luz e o sistema magnético e elétrico absoluto de unidades, bem como a relação entre o índice de refração e a constante dielétrica, a relação quantitativa entre o coeficiente e o reflexo de um corpo e sua condutividade metálica – tudo isso era como uma revelação. Sem contar a transição para uma teoria de campo, isto é, a expressão das leis elementares através de equações diferenciais, faltava a Maxwell apenas um simples passo hipotético – a introdução da corrente elétrica de deslocamento no vácuo e na dieletricidade e seu efeito magnético, uma inovação quase pré-ordenada pelas propriedades formais das equações diferenciais. A esse respeito devo observar que a dupla Faraday-Maxwell assemelha-se intrinsecamente à dupla Galileu-Newton – os primeiros descobrindo as relações por intuição, e os segundos formulando essas relações corretamente e aplicando-as quantitativamente.

O que dificultava a compreensão exata da essência da teoria eletromagnética, naquela época, era um fato bastante peculiar. “Intensidades de campo” elétrico ou magnético e “deslocamentos” eram tratados como variáveis elementares, e o espaço vazio, como uma fase especial do corpo dielétrico. A *matéria* aparecia como portadora do campo e não do *espaço*. Isso implicava a noção de que a portadora do campo deveria ter velocidade; e isso, naturalmente, aplicava-se também ao “vácuo” (éter). A eletrodinâmica dos corpos em movimento, de Hertz, baseia-se inteiramente nessa atitude fundamental.

O grande mérito de H. A. Lorentz foi a introdução de uma mudança convincente. A princípio um campo existe, segundo Hertz, apenas no espaço vazio. A matéria – considerada como constituída de átomos – é a única sede das cargas elétricas; entre as partículas

materiais há um espaço vazio, a sede do campo eletromagnético, produzido pela posição e velocidade dos pontos de carga localizados nas partículas da matéria. O comportamento dielétrico, condutividade etc., são determinados exclusivamente pelo tipo de conexões mecânicas existentes entre as partículas que constituem os corpos. As cargas das partículas criam o campo, o qual, por seu lado, exerce força sobre as cargas das partículas, determinando assim o movimento destas últimas, de acordo com a lei do movimento, de Newton. Se compararmos este sistema com o de Newton veremos que a diferença consiste no seguinte: a ação a distância é substituída pelo campo, o qual descreve também a radiação. A gravitação geralmente não é considerada, por ser relativamente pequena; entretanto, sua inclusão é sempre possível com o enriquecimento do campo, ou seja, com a expansão das leis de campo, de Maxwell. Os físicos da geração atual consideram a solução de Lorentz como a única possível; naquela época, entretanto, foi um passo surpreendente e audacioso, sem o qual não teria sido possível o desenvolvimento posterior.

Estudando criticamente esta fase do desenvolvimento da teoria, chama-nos a atenção o dualismo contido no fato de que o ponto material, segundo Newton, e o campo, como um contínuo, são usados como conceitos elementares, um ao lado do outro. A energia cinética e a energia de campo aparecem como coisas essencialmente diferentes. Esse dualismo torna-se mais insatisfatório quando o campo magnético de uma carga elétrica em movimento representa, segundo Maxwell, a inércia. Por que não a *totalidade* da inércia? Nesse caso, sobraria apenas a energia do campo, e a partícula seria apenas uma entidade com densidade especificamente alta de energia de campo. Nesse caso, seria possível deduzir o conceito de ponto de massa, com as equações de movimento das partículas, a partir das equações do campo – o dualismo seria assim removido.

H. A. Lorentz sabia disso muito bem. Entretanto, as equações de Maxwell não permitiam a derivação do equilíbrio da eletricidade que forma uma partícula. Somente equações diferentes, não-lineares, poderiam realizar isso. Não existia, porém, nenhum método para descobrir essas equações do campo sem descambar para uma

arbitrariedade arriscada. De qualquer modo, podia-se acreditar que, gradativamente, seria possível chegar a uma base fundamental nova e segura para toda a física, percorrendo o caminho iniciado com tanto sucesso por Faraday e Maxwell.

Assim, a revolução começada com a introdução do campo não terminara ainda. Então, no início do século, independente do que foi discutido acima, houve uma segunda crise fundamental, cuja gravidade imediatamente se reconheceu, devido às investigações de Max Planck sobre a radiação do calor (1900). A história desse acontecimento é notável, especialmente nessa primeira fase, por não ter sido influenciada de modo algum por descobertas casuais de natureza experimental.

No estudo da termodinâmica, Kirchhoff concluíra que a densidade da energia e a composição espectral da radiação numa cavidade com paredes impermeáveis e temperatura T deviam ser independentes da natureza das paredes. Ou seja, a densidade monocromática da radiação $\mathcal{E}$ é uma função universal da frequência e da temperatura absoluta T . Assim, surgiu o interessante problema de determinar esta função $\mathcal{E}(\nu, T)$. O que se poderia concluir teoricamente dessa função? Segundo a teoria de Maxwell, a radiação devia exercer pressão sobre as paredes, pressão essa determinada pela densidade total da energia. A partir desse ponto, Boltzmann concluiu, por meio da termodinâmica pura, que toda a densidade de energia da radiação ($\int \mathcal{E} d\nu$) é proporcional a T^4 . Desse modo encontrou a justificativa teórica de uma lei descoberta empiricamente por Stefan; isto é, fez a conexão entre a lei empírica e a base da teoria de Maxwell. Depois disso, por meio de uma engenhosa consideração termodinâmica, que usava também a teoria de Maxwell, W. Wien descobriu que a função universal $\mathcal{E}$ das duas variantes ν e T devia ser representada do seguinte modo:

$$\mathcal{E}, \nu^3 f\left(\frac{\nu}{T}\right)$$

onde $f(\nu/T)$ é uma função universal de uma variante ν/T . Evidentemente, a determinação teórica dessa função universal f não era fundamentalmente importante – essa a tarefa enfrentada por Planck. Medidas precisas tinham levado a uma determinação empírica bastante acurada da função f . Tomando como base essas medidas empíricas, consegui a princípio encontrar uma definição que representava perfeitamente as medidas:

$$\exists 0 \frac{81h\nu^3}{c^3} \frac{1}{\exp h\nu / kT. / 1},$$

onde h e k são duas constantes universais, sendo que a primeira levou à teoria dos quanta. Devido ao seu denominador, essa fórmula parece um tanto estranha. Seria possível derivá-la teoricamente? Na verdade, Planck encontrou uma derivada cujas imperfeições a princípio passaram despercebidas, o que veio a ser uma vantagem para o desenvolvimento da física. Se a sua fórmula estivesse correta, permitiria, com o auxílio da teoria de Maxwell, o cálculo da energia média E de um oscilador semimonocromático, dentro do campo da radiação:

$$E 0 \frac{h\nu}{\exp h\nu / kT. / 1}.$$

Planck preferiu tentar o cálculo teórico desta última grandeza. A termodinâmica e a teoria de Maxwell não foram de grande utilidade para esse empreendimento. A expressão tinha aspecto bastante promissor. Para altas temperaturas (com ν fixo) teremos a expressão:

$$E 0 kT.$$

É a mesma expressão obtida na teoria cinética dos gases para a energia média de um ponto material capaz de oscilar elasticamente em uma dimensão. Pois na teoria cinética dos gases teremos:

$$E = \frac{1}{2} N k T,$$

onde R representa o gás constante, e N o número de moléculas por mol, de cuja constante podemos computar o tamanho absoluto do átomo. Equacionando estas duas expressões, teremos

$$N = R / k.$$

A constante da fórmula de Planck, portanto, nos dá exatamente o peso correto do átomo. O valor numérico concorda satisfatoriamente com as determinações de N por meio da teoria cinética dos gases, embora esta última não seja muito exata.

Foi um grande sucesso, reconhecido por Planck. Mas havia um grave inconveniente, que felizmente ele não notou a princípio. Pois as mesmas considerações exigem que a relação $E = kT$ seja válida também para baixas temperaturas. Porém nesse caso a fórmula de Planck ficaria sem valor, bem como a constante h . Portanto, a conclusão correta da teoria existente seria: ou a energia cinética média do oscilador é dada incorretamente na teoria dos gases, o que implicaria a refutação da mecânica [estatística], ou a energia média do oscilador deriva-se incorretamente da teoria de Maxwell, o que implicaria a refutação desta última. Nessas circunstâncias, o mais provável é que as duas teorias sejam corretas apenas no limite, e falsas em tudo o mais; na verdade, essa é a situação real, como veremos a seguir. Se Planck tivesse chegado a essa conclusão, provavelmente não teria feito a grande descoberta, pois o raciocínio dedutivo puro não teria fundamento.

Voltando ao raciocínio de Planck. Baseado na teoria cinética dos gases, Boltzmann descobriu que, à exceção de um fator constan-

te, a entropia era igual ao logaritmo da “probabilidade” do estado em questão. De acordo com essa indicação, reconheceu a natureza dos processos que, por definição termodinâmica, são “irreversíveis”. Sob o ponto de vista mecânico molecular, entretanto, todos os processos são reversíveis. Se chamarmos a um estado, definido nos termos da teoria molecular, de estado microscopicamente descrito, ou microestado, e o estado, definido nos termos da termodinâmica, de macroestado, então um número imenso de estados (Z) pertence à condição macroscópica. Nesse caso, Z é a medida da probabilidade de um macroestado determinado. Esta idéia, aparentemente, é da maior importância, especialmente porque sua aplicabilidade não se limita a uma descrição microscópica com base na mecânica. Planck, reconhecendo esse fato, aplicou o princípio de Boltzmann a um sistema formado por vários ressoadores da mesma frequência ν . O estado macroscópico é dado pela energia total da oscilação de todos os ressoadores; o microestado é dado pela fixação da energia (instantânea) de cada ressoador individualmente. Para expressar o número de microestados pertencentes a um macroestado em número finito, ele [Planck] dividiu a energia total em um número grande mas finito de elementos idênticos de energia 2 e perguntou: de quantos modos podem esses elementos de energia ser divididos entre os ressoadores? O logaritmo desse número, então, nos dá a entropia e assim (por meio da termodinâmica), a temperatura do sistema. Planck obteve a fórmula da radiação determinando a grandeza dos elementos da energia 2 como $2 = h\nu$. O elemento decisivo desse processo é o fato de o resultado depender de darmos a um valor finito definido, isto é, não chegar ao limite $2 = 0$. Este raciocínio não evidencia o fato de que ele é contrário às bases da mecânica e da eletrodinâmica, das quais a derivação depende. Na verdade, a derivação pressupõe, implicitamente, que a energia pode ser absorvida e emitida pelo ressoador individual apenas em quanta de grandeza $h\nu$, isto é, que a energia de uma estrutura mecânica capaz de oscilações e a energia de radiação só podem ser transportadas nesses quanta – contrariando as leis da mecânica e da eletrodinâmica. A contradição com a dinâmica é fundamental, enquanto a contradição com a eletrodinâmica pode ser menos fundamental. Pois

a expressão para a densidade de energia de irradiação, embora *compat vel* com as equações de Maxwell, não é uma consequência necessária dessas equações. A prova de que essa expressão nos fornece valores médios importantes está no fato de que a lei de Stefan-Boltzmann e a lei de Wien, que se baseia nela, concordam com a experiência.

Tudo isso me pareceu evidente logo depois da publicação do trabalho básico de Planck; assim, embora sem um substituto para a mecânica clássica, podia ver quais seriam as consequências dessa lei da radiação da temperatura sobre o efeito fotoelétrico e sobre outros fenômenos relacionados com a transformação da radiação da energia, bem como sobre o calor específico dos corpos (especialmente) sólidos. Contudo, todas as minhas tentativas para adaptar os fundamentos teóricos da física a este [novo tipo de] conhecimento falharam completamente. Era como se subitamente nos faltasse o chão sob os pés, como se não existisse nenhuma base firme para construir. O fato de essa insegurança e contradição da fundação terem sido suficientes para levar Bohr, um homem de instinto acurado e sensibilidade privilegiada, a descobrir as leis principais das linhas espectrais e das camadas esféricas dos átomos e seu significado para a química, pareceu-me então um milagre – e até hoje me parece. É a mais alta forma de musicalidade na esfera do pensamento.

Nessa época, não concentrava meu interesse nos detalhes das consequências dos resultados obtidos por Planck, por mais importantes que fossem. Minha pergunta principal era: qual a conclusão geral a que se pode chegar partindo da fórmula da radiação, no que se refere à estrutura da radiação e, de um modo mais geral, no que se refere aos fundamentos eletromagnéticos da física? Antes de prosseguir neste assunto, devo mencionar brevemente algumas investigações relacionadas ao movimento browniano e a objetos relacionados a ele (fenômenos de flutuação), que repousam essencialmente na mecânica molecular clássica. Sem conhecer os estudos de Boltzmann e Gibbs, publicados anteriormente e que na verdade exauriram o assunto, desenvolvi a teoria da mecânica estatística e a teoria cinético-molecular da termodinâmica, que se baseia nela. Meu objetivo principal era encontrar fatos que asseguerrassem, da melhor forma

possível, a existência de átomos de tamanhos finitos definidos. Descobri então que, segundo a teoria atomística, deveria haver um movimento de partículas microscópicas em suspensão capaz de ser observado, sem saber que as observações sobre o movimento browniano eram há muito conhecidas. A derivação mais simples baseava-se no seguinte raciocínio. Se a teoria cinética molecular é essencialmente correta, uma suspensão de partículas visíveis deve possuir o mesmo tipo de pressão osmótica, de acordo com a lei que define os gases como uma solução de moléculas. A pressão osmótica depende da grandeza real das moléculas, isto é, do número de moléculas em um equivalente-grama, ou mol. Se a densidade da suspensão não for homogênea, a pressão osmótica também não será homogênea e provocará uma difusão compensatória, que pode ser calculada pela mobilidade conhecida das partículas. Essa difusão, por outro lado, pode também ser considerada como o resultado do deslocamento ao acaso – originalmente de grandeza desconhecida – das partículas em suspensão, devido à agitação térmica. Comparando as quantidades obtidas pela corrente de difusão dos dois tipos de raciocínio, obtém-se quantitativamente a lei estatística para esses deslocamentos, isto é, a lei do movimento browniano. A concordância dessas considerações com a experiência e com a determinação do verdadeiro tamanho molecular, de Planck, a partir da lei da radiação (para altas temperaturas) convenceu os mais céticos, que eram em grande número à época (Ostwald, Mach), da realidade do átomo. A hostilidade desses estudiosos para com a teoria atômica pode ser atribuída sem dúvida à sua atitude filosófica positivista. É um exemplo interessante de que mesmo os estudiosos de espírito audacioso e instinto apurado podem ter sua interpretação dos fatos prejudicada por preconceitos filosóficos. O preconceito – que não desapareceu até hoje – consiste em acreditar que os fatos podem e devem fornecer, por si mesmos, conhecimento científico, sem uma construção conceptual livre. Esse modo de pensar só é possível quando não se leva em conta a livre escolha dos conceitos, os quais, por meio dos resultados positivos e longo tempo de uso, parecem estar diretamente ligados ao material empírico.

O sucesso da teoria do movimento browniano demonstrou, mais uma vez de forma conclusiva, que a mecânica clássica sempre conduz a resultados fidedignos quando aplicada a movimentos nos quais as derivadas do maior tempo da velocidade são negligenciáveis. Essa afirmativa pode servir de base a um método relativamente direto, que nos fornece alguma informação sobre a constituição da radiação a partir da fórmula de Planck. Pode-se argumentar que, em um espaço ocupado pela radiação, um espelho de movimento livre (vertical ao seu plano), com reflexo semimonocromático, sofreria uma espécie de movimento browniano, cuja média de energia cinética seria igual a $\frac{1}{2} (R/N) T$ (sendo R = gás constante para uma molécula-grama, N = número de moléculas por mol, T = temperatura absoluta). Se a radiação não estiver sujeita a flutuações locais, o espelho gradualmente chegará à posição de repouso, pois, devido ao seu movimento, reflete mais radiação na parte da frente do que na de trás. Entretanto, o espelho sofrerá flutuações ao acaso, devidas à pressão exercida sobre ele, porque os grupos de ondas que constituem a radiação interferem entre si. Esse fato pode ser conferido pela teoria de Maxwell. Portanto, o cálculo demonstra que essas variações de pressão (especialmente no caso das densidades de pequena radiação) não são suficientes para transmitir ao espelho a energia cinética média $\frac{1}{2} (R/N) T$. Para chegar a esse resultado é preciso presumir a existência de um segundo tipo de variação de pressão, não derivada da teoria de Maxwell, correspondente ao pressuposto de que a energia de radiação consiste em um conjunto puntiforme de quanta, indivisível e localizado, de energia $h\nu$ [e de momentum $h\nu/c$, (c = velocidade da luz)], que são refletidos de modo indivisível. Essa abordagem do problema provou de modo drástico e direto que se deve atribuir um tipo de realidade imediata aos quanta de Planck; que, conseqüentemente, a radiação deve possuir uma espécie de estrutura molecular no que se refere à sua energia, o que naturalmente contradiz a teoria de Maxwell. As considerações sobre radiação, baseadas diretamente na relação de probabilidade da entropia, de Boltzmann (probabilidade elevada a igual freqüência temporal estatística), levam ao mesmo resultado. Essa dualidade na natureza da radiação (e dos corpúsculos materiais)

é uma das propriedades principais da realidade, que tem sido interpretada pela mecânica dos quanta de forma engenhosa e com surpreendente sucesso. Essa interpretação, tida como basicamente definitiva por quase todos os físicos contemporâneos, parece-me apenas um expediente temporário; mais adiante farei algumas observações a esse respeito.

Esse tipo de raciocínio levou-me a concluir, há muito tempo, pouco depois de 1900, ou seja, depois do trabalho pioneiro de Planck, que nem a mecânica e nem a eletrodinâmica podiam (exceto nos casos limites) reivindicar validade exata. Gradualmente perdi a esperança de descobrir as leis verdadeiras através dos esforços construtivos, baseados em fatos conhecidos. Quanto mais me dedicava a esse objetivo, mais me convencia de que só a descoberta de um princípio formal universal poderia levar a resultados seguros e positivos. O exemplo que tinha ante meus olhos era a termodinâmica. O princípio geral era dado pelo teorema: segundo as leis da natureza, é impossível construir um *perpetuum mobile* (do primeiro e do segundo tipo). Sendo assim, como seria possível encontrar esse princípio universal? Após dez anos de estudo, o princípio surgiu, resultando de um paradoxo com o qual me defrontara quando tinha dezesseis anos: se um raio luminoso for perseguido a uma velocidade c (velocidade da luz no vácuo), observamos esse raio de luz como um campo eletromagnético em repouso, embora com oscilação espacial. Entretanto, aparentemente não existe tal coisa, quer com base na experiência, quer de acordo com as equações de Maxwell. Desde o início, tive a intuição clara de que, segundo o ponto de vista desse observador, tudo devia acontecer de acordo com as mesmas leis aplicáveis a um observador que estivesse em repouso em relação à terra. Pois, como poderia o primeiro observador saber ou determinar que está em estado de movimento rápido uniforme?

Vemos nesse paradoxo o germe da teoria da relatividade restrita. Hoje todos sabem que as tentativas de esclarecer satisfatoriamente esse paradoxo estariam condenadas ao fracasso

enquanto o axioma do caráter absoluto do tempo, ou da simultaneidade, estivesse enraizado no inconsciente. A compreensão do axioma e do seu caráter arbitrário é o ponto essencial para a solução do problema. O tipo de raciocínio crítico necessário para a descoberta desse ponto central foi, no meu caso, enriquecido especialmente com a leitura das obras filosóficas de David Hume e de Ernst Mach.

Era preciso compreender o significado exato, na física, das coordenadas espaciais e do tempo de fixação de um acontecimento. A interpretação física das coordenadas espaciais pressupunha um corpo rígido como referência, o qual deveria estar em estado de movimento mais ou menos definido (sistema de inércia). Em um dado sistema de inércia, as coordenadas denotam os resultados de certas medidas feitas com uma barra graduada rígida (estacionária). (Devemos ter em mente o fato de que o pressuposto da existência, em tese, das barras rígidas é sugerido pela experiência aproximada, mas, em princípio, arbitrária.) Com essa interpretação das coordenadas espaciais, a questão da validade da geometria euclidiana passa a ser um problema da física.

Portanto, se tentarmos interpretar o tempo de um acontecimento analogamente, precisamos de um meio para medir a diferença de tempo (um processo periódico determinado internamente e realizado por um sistema de extensão espacial suficientemente pequena). Um relógio em repouso, em relação ao sistema da inércia, define o tempo local. Os tempos locais de todos os pontos do espaço tomados em conjunto constituem o "tempo", que pertence ao sistema de inércia dado, se for fornecido um meio de "acertar" esses relógios entre si. Vemos que, *a priori*, não é necessário que os "tempos", assim definidos em diferentes sistemas de inércia, concordem entre si. Essa observação teria sido feita há mais tempo se pela experiência prática a luz não apresentasse (devido ao grande valor de c) os meios para fixar uma simultaneidade absoluta.

Os pressupostos da existência (em princípio) das barras graduadas (ideais ou perfeitas) e de relógios não são independentes entre si; um sinal luminoso refletido entre as extremidades de uma

barra rígida constitui o relógio ideal, desde que o postulado da constância da velocidade da luz não conduza a contradições.

O paradoxo acima pode, então, ser formulado da seguinte maneira. De acordo com as regras da conexão, usadas na física clássica, entre as coordenadas espaciais e o tempo dos acontecimentos na transição de um sistema de inércia para outro, os dois pressupostos

- (1) a constância da velocidade da luz
- (2) a independência das leis (especialmente da lei da constância da velocidade da luz) na escolha do sistema de inércia (princípio da relatividade restrita)

são mutuamente compatíveis (apesar de tomados separadamente, com base na experiência).

O principal ponto de partida para a teoria da relatividade restrita é o seguinte: os pressupostos (1) e (2) serão compatíveis se relações de um novo tipo (“transformação de Lorentz”) forem postuladas para a conversão das coordenadas e dos tempos dos acontecimentos. Dada a interpretação física das coordenadas e do tempo, isto não representa apenas um passo convencional, mas implica certas hipóteses relativas ao comportamento real dos relógios e das barras graduadas em movimento, que podem ser confirmadas ou negadas pela experiência.

O princípio universal da teoria da relatividade restrita está contido no postulado: as leis da física são invariantes em relação às transformações de Lorentz (pela transição de um sistema de inércia para qualquer outro sistema de inércia escolhido arbitrariamente). Este é um princípio restritivo para as leis naturais, comparável ao princípio restritivo da não-existência do *perpetuum mobile* que é a base da termodinâmica.

Uma observação sobre a relação da teoria do “espaço quadridimensional”. Acredita-se erroneamente que a quadridimensionalidade do *continuum* físico tenha sido descoberta, até certo ponto, ou pelo menos introduzida pela teoria da relatividade restrita. Tal não se deu. A mecânica clássica também se baseia no

continuum quadridimensional do espaço e do tempo. Mas, no *continuum* quadridimensional da física clássica, os subespaços com valor de tempo constante têm uma realidade absoluta, independente da escolha da estrutura de referência. Por isso, o *continuum* quadridimensional divide-se naturalmente em um tridimensional e um unidimensional (tempo), de modo que o conceito de quadridimensional não se impõe como uma *necessidade*. A teoria da relatividade restrita, por sua vez, cria uma dependência formal entre o modo pelo qual as coordenadas do espaço, de um lado, e as coordenadas de tempo, de outro, se incorporam às leis naturais.

A importante contribuição de Minkowski à teoria é a seguinte: antes de sua investigação, era necessário confirmar uma transformação de Lorentz em uma lei para testar sua invariância sob essas transformações; porém, ele conseguiu introduzir um formalismo segundo o qual a forma matemática da lei *garante* sua invariância sob as transformações de Lorentz. Criando o cálculo tensor quadridimensional, obteve o mesmo resultado para o espaço quadridimensional obtido pelo cálculo vetorial comum para as dimensões espaciais tridimensionais. Provou também que a transformação de Lorentz (com exceção de um signo algébrico diferente, devido ao caráter especial do tempo) não é mais que a rotação do sistema de coordenadas no espaço quadridimensional.

Em primeiro lugar, uma observação crítica sobre a teoria descrita acima. É digno de nota o fato de a teoria introduzir (exceto para o espaço quadridimensional) dois tipos de coisas físicas, isto é (1) barras graduadas e relógios, e (2) todas as outras coisas, como, por exemplo, o campo eletromagnético, o ponto material etc. Isto é, de certa forma, inconsistente; estritamente falando, as barras graduadas e os relógios deveriam aparecer como soluções das equações básicas (objetos que consistem em configurações atômicas móveis) e não, como era o caso, como entidades teoricamente auto-suficientes. Esse procedimento se justifica, entretanto, por ter sido evidente, desde o princípio, que os postulados da teoria não são bastante fortes para que sejam deduzidas deles equações dos acontecimentos físicos, suficientemente completas e suficientemente livres da arbitrariedade para que possam ser usadas como base da teoria das barras graduadas

e dos relógios. Se não se desejava desenvolver uma interpretação física das coordenadas em geral (algo que por si só seria impossível), o melhor seria permitir essa inconsistência – com a obrigação, entretanto, de eliminá-la num estágio mais avançado da teoria. Porém, não se deve legitimar que as distâncias são entidades físicas de um tipo especial, intrinsecamente diferentes das outras variáveis físicas (“reduzir a física à geometria” etc.).

Estudaremos agora as elucidações de natureza definitiva que a física deve à teoria da relatividade restrita.

(1) Não existe o que se chama de simultaneidade dos acontecimentos distantes; portanto, não existe também a ação imediata a distância, no sentido da mecânica de Newton. Embora a introdução das ações a distância, que se propagam à velocidade da luz, continue praticável de acordo com essa teoria, ela não parece natural, pois nessa teoria não poderia haver uma expressão razoável para o princípio da conservação da energia. Portanto, parece inevitável que a realidade física seja descrita em termos de funções contínuas no espaço. Logo, o ponto material dificilmente pode ser tomado como conceito básico da teoria.

(2) Os princípios da conservação do *momentum* linear e da energia fundem-se em um único princípio. A massa inerte de um sistema isolado é idêntica à sua energia, eliminando assim a massa como conceito independente.

Observação. A velocidade da luz c é uma das quantidades que ocorrem nas equações físicas como uma "constante universal". Porém, se for introduzida como unidade de tempo, em vez do segundo, o tempo gasto pela luz para percorrer 1 cm, c não ocorrerá nas equações. Sob esse ponto de vista, pode-se dizer que a constante c é apenas uma constante universal *aparente*.

É óbvio e geralmente aceito o fato de que se pode eliminar mais duas constantes universais da física introduzindo, no lugar do grama e do centímetro, unidades “naturais” adequadamente escolhidas (por exemplo, a massa e o raio do elétron).

Se considerarmos que o processo foi realizado, então apenas as constantes “não-dimensionais” poderão ocorrer nas equações básicas da física. A esse respeito gostaria de colocar uma proposição que

atualmente só pode ser baseada na fé e na simplicidade, isto é, na inteligibilidade da natureza: não existem constantes *arbitrárias* desse tipo, ou seja, a natureza permite, por sua constituição, que sejam enunciadas leis tão fortemente determinadas que nelas ocorram somente constantes completa e racionalmente determinadas (portanto, não constantes cujo valor numérico possa ser mudado sem destruir a teoria).

A teoria da relatividade restrita se originou das equações do campo eletromagnético de Maxwell. Por outro lado, este último só pode ser compreendido de modo satisfatório por meio da teoria da relatividade restrita. As equações de Maxwell são as mais simples equações de campo invariável de Lorentz, que podem ser postuladas para um tensor anti-simétrico derivado de um campo vetor. Isto seria suficiente, se não soubéssemos, por meio dos fenômenos dos quanta, que a teoria de Maxwell não faz justiça às propriedades energéticas da radiação. Porém nem mesmo a teoria da relatividade restrita oferece base adequada quanto à maneira pela qual a teoria de Maxwell deveria ser modificada de modo natural. Além disso, à indagação de Mach, “por que distinguir os sistemas de inércia entre todos os outros sistemas de coordenadas?”, essa teoria não oferece resposta.

O fato de ser a teoria da relatividade restrita apenas o primeiro passo de um desenvolvimento necessário só se tornou evidente para mim quando procurei representar a gravitação na estrutura dessa teoria. Na mecânica clássica, interpretada em termos de campo, o potencial da gravitação aparece como um campo *escalar* (a possibilidade teórica mais simples de um campo com um componente único). Essa teoria escalar do campo gravitacional pode ser facilmente transformada em invariante sob o grupo das transformações de Lorentz. Por isso, o programa seguinte parece natural: o campo físico total consiste em um campo escalar (gravitação) e um campo vetorial (campo eletromagnético); estudos posteriores podem tornar necessária a introdução de campos mais

complexos; porém, para começar, não precisamos nos preocupar com isso.

A possibilidade de realização desse programa, contudo, era duvidosa desde o princípio, porque a teoria devia ser uma combinação do seguinte:

- (1) As considerações sobre a teoria da relatividade restrita demonstravam claramente que a massa inerte de um sistema físico aumenta com a energia total (logo, aumenta com a energia cinética, por exemplo).
- (2) Experiências precisas (especialmente as experiências sobre o equilíbrio da torção, de Eötvös) provaram empiricamente, com grande exatidão, que a massa gravitacional de um corpo é igual à sua massa inerte.

Segue-se, por (1) e (2), que *o peso* de um sistema depende, de modo perfeitamente conhecido, da sua energia total. Se a teoria não chegasse a essa conclusão, ou não pudesse chegar a ela naturalmente, deveria ser rejeitada. A condição expressa-se com mais naturalidade da seguinte maneira: a aceleração de um sistema de queda livre no campo gravitacional é independente da natureza do sistema de queda (especialmente do seu conteúdo de energia).

Contudo, dentro da estrutura do programa traçado, este simples estado de coisas não podia ser representado de modo satisfatório, pelo menos não de um modo natural. Isso me convenceu de que dentro da estrutura da teoria da relatividade restrita não existe lugar para uma teoria adequada da gravitação.

Ocorreu-me, então, o seguinte: a igualdade entre a massa inerte e a massa gravitacional, ou seja, a independência da aceleração gravitacional da natureza da substância em queda, pode ser expressa da seguinte forma: Em um campo gravitacional (de pequena extensão espacial), os objetos comportam-se do mesmo modo que no espaço livre de gravitação se introduzirmos nele, em vez de um “sistema de

inércia”, uma estrutura de referência com aceleração relativa ao primeiro.

Se for interpretado o comportamento de um corpo em relação a estrutura de referência como provocado por um campo gravitacional “real” (não apenas aparente), é possível considerar esta estrutura como um “sistema de inércia”, muito mais justificadamente do que o sistema de referência original.

Assim, se considerarmos campos gravitacionais difusos, não restritos *a priori* por condições de limites espaciais, como fisicamente possíveis, então o conceito do “sistema de inércia” torna-se completamente vazio. O conceito de “aceleração relativa ao espaço” perde todo o sentido, bem como o princípio de inércia do paradoxo de Mach.

A igualdade entre a massa inerte e a gravitacional leva assim, naturalmente, à conclusão de que o postulado básico da teoria da relatividade restrita (a invariância das leis sob as transformações de Lorentz) é muito limitado, isto é, a invariância das leis deve ser também postulada em relação às transformações *n^o-lineares* das coordenadas no contínuo quadridimensional.

Isso tudo aconteceu em 1908. Por que foram necessários outros 7 anos para a construção da teoria da relatividade geral? A principal razão é o fato de que não é fácil se libertar da idéia de que as coordenadas precisam ter um significado métrico direto. A transformação ocorreu mais ou menos do seguinte modo.

Começamos com um espaço vazio, sem campo, como é o caso – em relação ao sistema de inércia – na teoria da relatividade restrita, como a situação física mais simples que se pode imaginar. Se pensarmos agora em um sistema não-inerte introduzido no pressuposto de que o novo sistema é uniformemente acelerado contra o sistema da inércia (numa definição tridimensional) em uma direção (convenientemente definida), então existe, com referência a esse sistema, um campo gravitacional paralelo estático. O sistema de referência pode ser rígido, euclidiano, nas suas propriedades métricas tridimensionais. Mas o tempo em que o campo aparece como estático *n^o* é medido por relógios estacionários *igualmente constituídos*. Com esse exemplo, reconhecemos que o significado métrico ime-

diato das coordenadas desaparece quando se admitem as transformações não-lineares das coordenadas. Contudo, essa admissão é *obrigatória*, se quisermos fazer justiça à igualdade da massa gravitacional e da massa inerte, de acordo com as bases da teoria, e se tivermos como objetivo superar o paradoxo de Mach em relação aos sistemas de inércia.

Portanto, se nos libertarmos do conceito que atribui um significado métrico imediato às coordenadas (diferenças de coordenadas = comprimento mensurável, ou tempo) devemos considerar como equivalente todo sistema de coordenada que pode ser criado pelas transformações contínuas das mesmas.

A teoria da relatividade geral, portanto, parte do seguinte princípio: as leis naturais devem ser expressas por equações que são covariantes sob o grupo de transformações contínuas das coordenadas. Este grupo substitui o grupo das transformações de Lorentz, da teoria da relatividade restrita, que passa a ser um subgrupo do primeiro.

Evidentemente, este postulado não é suficiente como ponto de partida para a derivação das equações básicas da física. Para começar, pode-se negar que o postulado, por si só, represente uma restrição real às leis da física; pois é sempre possível reformular uma lei desenvolvida a princípio apenas para determinados sistemas de coordenadas, de modo que a nova formulação venha a ser uma variante geral formal. Além disso, é evidente que pode ser formulado um número infinitamente grande de leis de campo, que tenham essa propriedade de covariante. O significado eminentemente heurístico do princípio da relatividade geral é que a mesma nos leva à procura desses sistemas de equações que sejam, na sua *covariância geral*, a formulação das *mais simples possíveis*; entre essas, devemos procurar as equações de campo do espaço físico. Campos que podem ser transformados entre si por meio dessas transformações descrevem a mesma situação real.

A questão principal para o pesquisador desse campo é a seguinte: qual o tipo matemático das variáveis (funções das coordenadas) que permite a expressão das propriedades físicas do

espaço (“estrutura”)? E só depois disso, perguntar: quais as equações satisfeitas por essas variáveis?

Não temos ainda hoje as respostas certas para essas questões. O caminho escolhido para a primeira formulação da teoria da relatividade geral pode ser descrito da seguinte maneira. Embora não saibamos qual o tipo de variáveis de campo (estrutura) que irá caracterizar o espaço físico, conhecemos com certeza um caso especial: o caso do espaço de “campo livre” da teoria da relatividade restrita. Esse espaço caracteriza-se pelo fato de que, para um sistema de coordenadas adequadamente escolhido, a expressão

$$ds^2 = 0 \, dx_1^2 + 3 \, dx_2^2 + 3 \, dx_3^2 + dx_4^2 \quad (1)$$

pertencente a dois pontos vizinhos representa uma quantidade mensurável (quadrado da distância), tendo portanto significado físico real. Referida a um sistema arbitrário, essa quantidade é expressa do seguinte modo:

$$ds^2 = 0 \, g_{ik} dx_i dx_k \quad (2)$$

onde os índices vão de 1 a 4. O g_{ik} forma um tensor simétrico (real). Se após realizar a transformação no campo (1) as primeiras derivadas de g_{ik} não desaparecem em relação às coordenadas, existe um campo gravitacional com referência a esse sistema de coordenadas, no sentido da consideração exposta acima, mas de tipo muito especial. Graças à investigação de Riemann sobre os espaços métricos n -dimensionais, esse campo especial pode ser caracterizado invariantemente:

- (1) A curvatura-tensor de Riemann, $R_{ik/m}$, formada a partir de coeficientes do métrico (2), desaparece.
- (2) A trajetória de um ponto-massa em referência ao sistema de inércia [em relação ao qual (1) é válido] é uma linha reta, portanto um extremal (geodésico). Esta última afirmação,

entretanto, é uma caracterização da lei do movimento baseada em (2).

A lei *universal* do espaço físico deve ser uma generalização da lei que acabamos de descrever. Neste ponto, são pressupostos dois passos para a generalização:

- (a) o campo gravitacional puro
- (b) o campo geral (que incluirá também quantidades que de certa forma correspondem ao campo eletromagnético).

O caso (a) foi caracterizado pelo fato de que o campo pode ainda ser representado por um métrico Riemann (2), isto é, por um tensor simétrico, mas sem representação da forma (1) (a não ser em escala infinitesimal). Isto significa que, no caso (a), o tensor de Riemann não desaparece. Entretanto, é evidente que deve haver uma lei que represente a generalização (expansão) dessa lei. Se essa lei generalizada for também da segunda ordem de diferenciação e linear nas segundas derivadas, apenas a equação obtida por uma simples contração

$$0 \ 0 \ R_{kl} \ 0 \ g^{im} R_{iklm}$$

seria uma lei prospectiva do campo no caso (a). Parece natural, além disso, pressupor que, no caso (a), a linha geodésica representa ainda a lei de movimento do ponto material.

Naquela época pareceu-me inútil aventurar-me na tentativa de representar o campo total (b) e determinar leis de campo para o mesmo. Portanto, decidi criar uma estrutura formal preliminar para a representação de toda a realidade física; isto era necessário para a investigação, pelo menos preliminar, da eficácia da idéia básica da relatividade geral. O que fiz, do seguinte modo:

Na teoria de Newton, pode-se escrever a lei do campo de gravitação assim:

$$\nabla^2 \Phi = 0$$

(Φ = gravitação potencial), válida sempre que a densidade da matéria, ρ , desaparece. Em geral temos (equação de Poisson):

$$\nabla^2 \Phi = 4\pi k \rho \quad (\rho = \text{densidade da massa})$$

Na teoria relativista do campo gravitacional, R_{ik} toma o lugar de $\nabla^2 \Phi$. No lado direito, portanto, temos de substituir o ρ por um tensor. Uma vez que sabemos, segundo a teoria da relatividade restrita, que a massa (inerte) é igual à energia, devemos colocar do lado direito o tensor de densidade de energia – mais precisamente, de toda a densidade de energia que não pertence ao campo gravitacional puro. Chegamos assim à equação de campo:

$$R_{ik} - \frac{1}{2} g_{ik} R = -k T_{ik}.$$

O segundo membro do lado esquerdo é adicionado devido a considerações formais, pois o lado esquerdo é escrito de forma que sua divergência, no sentido do cálculo diferencial absoluto, desapareça identicamente. O lado direito é uma condensação formal de todas as coisas cuja compreensão, no sentido de uma teoria de campo, é ainda problemática. Nem por um momento, naturalmente, duvidei que essa formulação fosse apenas um recurso temporário, apenas uma expressão preliminar do princípio da relatividade geral de forma fechada. Pois era essencialmente *nada mais* do que uma teoria do campo gravitacional, isolada artificialmente do campo total, de estrutura ainda desconhecida.

Se existe algo na teoria delineada acima – exceto o postulado de invariância das equações, sob o grupo de transformações contínuas das coordenadas – que possa ser considerado definitivo, é a teoria do caso da limitação de um campo gravitacional puro e sua

relação com a estrutura métrica do espaço. Por esse motivo, falaremos a seguir apenas das equações do campo gravitacional puro.

A peculiaridade dessas equações reside, por um lado, na sua complicada estrutura, especialmente no seu caráter não-linear com respeito as variáveis de campo e suas derivadas e, por outro lado, na necessidade quase compulsória com a qual o grupo de transformações determina essa complexa lei de campo. Se tivéssemos parado na teoria da relatividade restrita, isto é, na invariância sob o grupo de Lorentz, a lei do campo $R_{ik} = 0$ permaneceria invariante também, dentro da estrutura desse grupo mais restrito. Mas, sob o ponto de vista do grupo mais restrito, não haveria motivo imediato para representar a gravitação por meio de uma estrutura tão complexa como a do tensor simétrico g_{ik} . Se, apesar disso, houvesse razões suficientes para tanto, surgiria um número imenso de leis de campo, a partir das quantidades g_{ik} , todas elas covariantes sob as transformações de Lorentz (mas não sob o grupo geral). Porém, mesmo que entre todas as leis invariantes de Lorentz uma delas tivesse sugerido exatamente a lei pertencente ao grupo mais vasto, ainda assim não chegaríamos ao nível de compreensão correspondente ao princípio da relatividade geral. Pois, sob o ponto de vista do grupo de Lorentz, duas soluções teriam de ser consideradas, incorretamente, como fisicamente diferentes, se pudessem ser transformadas uma na outra por meio de uma transformação não-linear das coordenadas, isto é, se do ponto de vista do grupo mais extenso, fossem apenas representações diferentes do mesmo campo.

Ainda uma observação geral sobre estrutura e grupo. É evidente que, de um modo geral, julgamos uma teoria mais próxima da perfeição quanto mais simples for a “estrutura” por ela postulada e quanto mais extenso o grupo em relação ao qual as equações do campo são invariantes. Vemos agora que esses dois *desiderata* se contrariam. Por exemplo: segundo a teoria da relatividade restrita (grupo de Lorentz), pode-se determinar uma lei covariante para a mais simples das estruturas (um campo escalar), ao passo que, na teoria da relatividade geral (grupo mais extenso das transformações contínuas das coordenadas), há uma lei de campo invariante apenas

para a estrutura mais complexa do tensor simétrico. Já apresentamos as razões *físicas* para o fato de que, na física, deve-se exigir a invariância sob o grupo mais extenso:³ sob um ponto de vista puramente matemático, não vejo necessidade de sacrificar a estrutura mais simples em favor da generalidade do grupo.

O grupo da relatividade geral é o primeiro a exigir que a mais simples lei invariante deixe de ser linear e homogênea nas variáveis de campo e nas suas derivadas. Isto é de importância fundamental pelo seguinte: se a lei de campo for linear (e homogênea), a soma de duas soluções será outra solução; é o que acontece, por exemplo, nas equações de campo de Maxwell para o vácuo. Nessa teoria é impossível deduzir das equações de campo apenas uma interação entre estruturas que representam, separadamente, soluções do sistema. Por isso, todas as teorias até agora precisavam, além das equações de campo, de equações especiais para o movimento dos corpos materiais sob a influência dos campos. É verdade que na teoria relativista da gravitação, a lei do movimento (linha geodésica) foi a princípio postulada independentemente, além da lei de campo. Entretanto, foi demonstrado que a lei do movimento não precisa (e não deve) ser assumida independentemente, mas está implicitamente contida na lei do campo gravitacional.

A essência desta situação realmente complexa pode ser vista da seguinte forma: um único ponto material em repouso será representado por um campo gravitacional que é finito e regular em todas as partes, exceto onde está localizado o ponto material; aí o campo não tem singularidade. Entretanto, se completarmos o campo pertencente a dois pontos materiais em repouso, integrando as equações do campo, esse campo terá, além das singularidades na posição dos pontos materiais, uma curva de pontos singulares ligando os dois pontos. Entretanto, é possível estipular um movimento dos pontos materiais, de modo que o campo gravitacional determinado por eles não se torne singular em nenhuma parte, exceto

³ Permanecer com o grupo menos extenso e ao mesmo tempo basear a teoria da relatividade da gravitação nas estruturas mais complexas [tensores] implica uma inconsequência ingênua. O peccado continua sendo peccado, mesmo quando cometido por homens respeitáveis.

nos pontos materiais. Esses são exatamente os conceitos descritos na primeira abordagem das leis de Newton. Podemos dizer, portanto: as massas se movem de tal forma que a solução das equações de campo não é singular em lugar nenhum, exceto nos pontos de massa. Essa propriedade das equações gravitacionais está intimamente ligada à sua não-linearidade, e esta, por sua vez, resulta do grupo mais extenso de transformações.

Naturalmente poderíamos objetar: se são permitidas singularidades nos locais dos pontos materiais, qual a justificativa para proibir a ocorrência de singularidades em outros lugares? A objeção seria justificada se as equações da gravitação fossem consideradas como equações do campo total. [Uma vez que esse não é o caso], entretanto, diríamos que o campo de uma partícula material deveria diferir tanto mais de um *campo gravitacional puro* quanto mais próximo se chegasse ao local da partícula. Se tivéssemos as equações do campo total, seríamos levados a exigir que as próprias partículas pudessem ser representadas como soluções das equações do campo completo, que são livres de irregularidades em todos os lugares. Só então a teoria da relatividade geral seria uma teoria *completa*.

Antes de abordar a questão da conclusão da teoria da relatividade geral, devo definir-me sobre a mais eficiente teoria física da nossa época, ou seja, a teoria estatística dos quanta, que tomou forma lógica consistente há mais ou menos vinte e cinco anos (Schrödinger, Heisenberg, Dirac, Born). Atualmente é a única teoria que permite uma compreensão unitária das experiências sobre o caráter dos quanta nos acontecimentos micromecânicos. Essa teoria e a teoria da relatividade são consideradas corretas em certo sentido, embora tenham falhado até agora todos os esforços para fundi-las num todo. Talvez por isso, entre os físicos teóricos contemporâneos, as opiniões variam tanto sobre como será encarada a fundação teórica da física no futuro. Será uma teoria de campo? Será, em essência, uma teoria estatística? Exporei brevemente a minha opinião a esse respeito.

A física é uma tentativa de compreensão conceptual da realidade, considerada como algo independente da observação. Nesse sentido, falamos de “realidade física”. Na física anterior à teoria dos

quanta, não havia dúvida quanto ao modo de compreender esse conceito. Na teoria de Newton a realidade era determinada por um ponto material no espaço e no tempo; na teoria de Maxwell, pelo campo no espaço e no tempo. Na teoria mecânica dos quanta a situação é menos clara. A pergunta: a função 4 da teoria dos quanta representa um fato real no mesmo sentido que um sistema material de pontos ou um campo eletromagnético? – hesitamos em responder com um simples “sim” ou “não”. Por quê? O que a função 4 (num tempo definido) formula é o seguinte: qual a probabilidade de encontrar uma quantidade física q (ou p) em um dado intervalo definido, se o tempo usado para medi-lo for t ? A probabilidade deve ser vista como uma determinável empírica e portanto, evidentemente, uma quantidade “real”, que deve ser determinada se se criar a função 4 muitas vezes e se for feita, de cada vez, uma medição q . O sistema individual respectivo possui esse valor q antes mesmo da medição? Não há resposta definitiva a essa questão na estrutura da teoria [existente], uma vez que a medição é um processo que implica um distúrbio finito do sistema, vindo do exterior; seria portanto concebível que o sistema obtivesse um valor numérico definido para q (ou p), valor numérico medido apenas através da sua própria medição. Para maior esclarecimento, suponhamos dois físicos A e B, que representam conceitos diferentes sobre a situação real descrita pela função 4.

- A. O sistema individual (antes da medição) tem um valor definido de q (ou p) para todas as variáveis do sistema, especificamente *aquele* valor determinado por uma medição dessa variável. Partindo desse conceito, ele dirá: a função 4 não é uma descrição completa do estado exato do sistema, mas apenas uma representação incompleta; expressa somente o que sabemos sobre o sistema por meio de medições prévias.
- B. O sistema individual (antes da medição) não tem valor definido para q (ou p). O valor medido é produzido pelo próprio ato de medir, coerente com a probabilidade própria da função 4. Partindo desse conceito, ele dirá (ou, pelo menos, poderá di-

zer): a função Ψ é uma descrição exhaustiva da situação real do sistema.

Agora, apresentamos a esses dois físicos o seguinte caso. Suponhamos um sistema que no tempo t da nossa observação consiste em dois sistemas componentes, S_1 e S_2 , que no momento estão separados espacialmente (no sentido da física clássica), com pequena interação entre ambos. O sistema total deverá ser descrito completamente em termos da mecânica dos quanta, por uma função Ψ conhecida, digamos Ψ_{12} . Todos os teóricos dos quanta concordam com o seguinte: se fizermos uma medição completa de S_1 , obteremos do resultado da medição e de Ψ_{12} uma função Ψ completamente definida, Ψ_2 do sistema S_2 . O caráter de Ψ_2 então depende do tipo de medição que for feita em S_1 .

Na minha opinião, pode-se falar do estado real do sistema parcial S_2 . Para começar, antes de realizar a medição de S_1 , sabemos menos sobre o seu estado real do que sabemos sobre o sistema descrito pela função Ψ . Mas acho que devemos insistir em um pressuposto sem qualificações: o estado real do sistema S_2 é independente de qualquer manipulação do sistema S_1 , que está espacialmente separado do primeiro. Porém, de acordo com o tipo de medição feito em S_1 , teremos um Ψ_2 muito diferente para o segundo sistema parcial ($\Psi_2, \Psi_2^1, \dots$). Entretanto, o estado real de S_2 deve ser independente do que acontece a S_1 . Logo, é possível encontrar (dependendo da escolha da medição feita em S_1) para o estado real de S_2 tipos diferentes da função Ψ . (Só se escapa a essa conclusão pressupondo que a medição de S_1 altera o estado real de S_2 [telepaticamente], ou negando completamente que entidades espacialmente separadas possuam estados reais independentes. As duas alternativas me parecem totalmente inaceitáveis.)

Se os físicos A e B aceitarem o raciocínio como válido, B terá de abandonar a sua posição, segundo a qual a função Ψ constitui uma descrição completa do estado real. Pois, nesse caso, seria impossível atribuir dois tipos diferentes de funções Ψ ao estado idêntico de S_2 .

O caráter estatístico dessa teoria origina-se necessariamente da insuficiência da descrição dos sistemas na mecânica dos quanta, e

não haveria motivo algum para pressupor que, no futuro, a física viesse a ter como fundamento básico a estatística. Em minha opinião, a teoria contemporânea dos quanta representa a melhor formulação do relacionamento, dados certos conceitos básicos fixos provindos quase todos da mecânica clássica. Entretanto, acredito que essa teoria não oferece um ponto de partida apropriado para o desenvolvimento futuro. Neste ponto, minhas expectativas desviam-se bastante das expectativas dos físicos contemporâneos. Estão convencidos de que é impossível explicar os aspectos essenciais dos fenômenos dos quanta (mudanças aparentemente descontínuas de um sistema e não determinadas no tempo, qualidades simultaneamente corpusculares e ondulatórias dos transportadores elementares de energia) com uma teoria que descreve o estado real das coisas [objetos] por meio de funções contínuas de espaço para as quais são válidas equações diferenciais. Sou também de opinião que desse modo não é possível compreender a estrutura atômica da matéria e da radiação. Esperam que sistemas de equações diferenciais, que podem ser considerados para essa teoria, não tenham de modo algum soluções regulares (livres de singularidades) em todas as partes do espaço quadridimensional. Acima de tudo, entretanto, acreditam que o caráter aparentemente descontínuo dos processos elementares só pode ser descrito por meio de uma teoria essencialmente estatística, na qual as mudanças descontínuas dos sistemas são explicadas pelas mudanças contínuas das probabilidades dos possíveis estados.

Todas essas observações parecem-me importantes. Mas o ponto central é, para mim, a questão: o que pode ser feito, com alguma esperança de sucesso, tendo em vista a situação atual da teoria física? Minhas esperanças estão nas experiências com a teoria da gravitação. Para mim, essas equações têm maiores probabilidades de nos fornecer informação *precisa* do que quaisquer outras equações da física. Tomemos, por exemplo, as equações de Maxwell para o espaço vazio, como comparação. São formulações que correspondem às nossas experiências com campos eletromagnéticos infinitamente fracos. Essa origem empírica é suficiente para determinar sua forma linear; entretanto, como já foi acentuado acima, as verdadeiras leis não podem ser lineares. As leis lineares satisfazem o princípio de

superposição para as suas soluções; assim, não contêm afirmações sobre a interação dos corpos elementares. As verdadeiras leis não podem ser lineares e não podem ser derivadas de leis lineares. Aprendi algo mais com a teoria da gravitação: nenhuma coleção de fatos empíricos, por mais abrangente que seja, pode levar a essas equações complexas. Uma teoria deve ser testada pela experiência, mas não é possível construir uma teoria partindo da experiência. Equações complexas como as do campo gravitacional só podem ser encontradas através da descoberta de uma condição simplesmente matemática, que determine as equações de forma completa, ou quase completa. Uma vez obtidas essas condições extremamente formais, basta um pequeno conhecimento dos fatos para se construir a teoria, no caso das equações da gravitação, a quadridimensionalidade e o tensor simétrico como expressão da estrutura do espaço, ao lado da invariância relativa à contínua transformação do grupo, determinam as equações quase completamente.

Nossa tarefa consiste em encontrar as equações de campo para o campo total. A estrutura ideal deve ser uma generalização do tensor simétrico. O grupo não deve ser menos extenso do que o das transformações contínuas das coordenadas. Todas as tentativas nesse sentido fracassaram. Cheguei a negligenciar um aumento evidente ou oculto no número de dimensões do espaço, um processo originalmente usado por Kaluza, cuja variante projetada tem partidários até hoje. Devemos nos limitar ao espaço quadridimensional e ao grupo das transformações contínuas e reais as coordenadas. Após muitos anos de procura infrutífera, considero a solução delineada a seguir como a mais satisfatória logicamente.

Em lugar do g_{ik} ($g_{ik} = g_{ki}$) simétrico, é introduzido o tensor g_{ik} não-simétrico. Esta quantidade é composta de uma parte simétrica, s_{ik} e de outra, assimétrica e puramente imaginária, a_{ik} , assim:

$$g_{ik} = s_{ik} + i a_{ik}$$

Sob o ponto de vista do grupo, a combinação de s com a é arbitrária, porque os tensores s e a individualmente possuem caráter tensor. Entretanto, acontece que esses g_{ik} (vistos como um todo)

desempenham um papel bastante análogo na construção do campo gravitacional puro.

Esta generalização da estrutura do espaço parece natural também sob o ponto de vista do nosso conhecimento de física, pois sabemos que o campo eletromagnético envolve um tensor anti-simétrico.

Para a teoria da gravitação é também essencial que seja possível formar, a partir do g_{ik} simétrico, a densidade escalar

$$\sqrt{|g_{ik}|},$$

bem como o tensor contravariante g^{ik} , de acordo com a definição

$$g_{ik} g^{il} = \delta_k^l \quad (\delta_k^l \text{ 0 tensor de Kronecker}).$$

Essas estruturas podem ser definidas em correspondência exata para o g_{ik} não-simétrico, incluindo as densidades do tensor.

Na teoria da gravitação é ainda essencial que, para um campo g_{ik} simétrico, possa ser definido um campo g_{ik}^l , o qual seja simétrico nos subscritos e que, considerado geometricamente, governe o deslocamento paralelo de um vetor. Analogamente, para o g_{ik} não-simétrico deve ser definido um campo não-simétrico g_{ik}^l , de acordo com a fórmula

$$g_{ik,l} - g_{sk} g_{il}^s - g_{is} g_{lk}^s = 0, \quad (A)$$

que concorda com a relação correspondente do g simétrico; naturalmente, é preciso prestar atenção à posição dos índices mais baixos em g e g^s .

Como na teoria com o g_{ik} simétrico, é possível formar uma curvatura R_{klm}^i a partir de 5, e a partir de uma curvatura contraída R_{kl} . Finalmente, usando um princípio variacional com (A), é possível encontrar equações de campo compatíveis:

$$g^{i\hat{s}}_{,is} = 0 \quad (g^{i\hat{k}} = 0 \quad \frac{1}{2}(g^{ik} / g^{ki}) \sqrt{|g^{ik}|}) \quad (B_1)$$

$$5^{s}_{i\hat{s}} = 0 \quad (5^{s}_{i\hat{s}} = 0 \quad \frac{1}{2}(5^{s}_{is} / 5^{s}_{si})) \quad (B_2)$$

$$R_{i\hat{k}} = 0 \quad (C_1)$$

$$R_{k\hat{l}^*m} = 3 R_{lm^*k} = 3 R_{mk^*l} = 0 \quad (C_2)$$

Cada uma das duas equações (B_1 e B_2) é uma consequência da outra, se (A) for satisfeita. R_{kl} denota a parte simétrica, $R_{kl}^{\wedge}$, a parte anti-simétrica de R_{kl} .⁴

Se a parte anti-simétrica de g_{ik} desaparecer, essas fórmulas reduzem-se a (A) e (C_1) – o caso do campo gravitacional puro.

Acredito que essas equações constituam a generalização mais natural das equações da gravitação.⁵ A prova da sua utilidade física é

⁴ Devido a limitações tipográficas, as notações usadas nas fórmulas desta página são ligeiramente diferentes daquelas usadas no original. Os sinais de circunflexo (^) e travessão (–) substituíram o circunflexo invertido e o subscrito.

⁵ A teoria aqui proposta tem, na minha opinião, grande probabilidade de ser válida, se for possível descobrir o caminho para uma descrição exhaustiva da realidade física, baseada no contínuo.

uma tarefa extremamente complexa, uma vez que meras aproximações não são suficientes.

O problema é: quais são as soluções dessas equações que são regulares em todos os lugares?

Esta exposição terá alcançado o seu propósito se conseguir mostrar como os esforços de toda uma vida se combinam e por que levam a determinadas expectativas.

A handwritten signature in black ink, reading "A. Einstein." with a horizontal line extending from the end of the name.

Institute of Advanced Study
Princeton, Nova Jersey
[ca. 1946]